The composite method: an improved method for stream-water solute load estimation

Brent T. Aulenbach* and Richard P. Hooper[†]

US Geological Survey, Peachtree Business Center, 3039 Amwiler Road, Suite 130, Atlanta, GA 30360-2824, USA

Abstract:

The composite method is an alternative method for estimating stream-water solute loads, combining aspects of two commonly used methods: the regression-model method (which is used by the composite method to predict variations in concentrations between collected samples) and a period-weighted approach (which is used by the composite method to apply the residual concentrations from the regression model over time). The extensive dataset collected at the outlet of the Panola Mountain Research Watershed (PMRW) near Atlanta, Georgia, USA, was used in data analyses for illustrative purposes. A bootstrap (subsampling) experiment (using the composite method and the PMRW dataset along with various fixed-interval and large storm sampling schemes) obtained load estimates for the 8-year study period with a magnitude of the bias of less than 1%, even for estimates that included the fewest number of samples. Precisions were always <2% on a study period and annual basis, and <2% precisions were obtained for quarterly and monthly time intervals for estimates that had better sampling. The bias and precision of composite-method load estimates varies depending on the variability in the regression-model residuals, how residuals systematically deviated from the regression model over time, sampling design, and the time interval of the load estimate. The regression-model method did not estimate loads precisely during shorter time intervals, from annually to monthly, because the model could not explain short-term patterns in the observed concentrations. Load estimates using the period-weighted approach typically are biased as a result of sampling distribution and are accurate only with extensive sampling. The formulation of the composite method facilitates exploration of patterns (trends) contained in the unmodelled portion of the load. Published in 2006 by John Wiley & Sons, Ltd.

KEY WORDS   stream-water solute loads; water quality; trends; sample design; bootstrap experiment

INTRODUCTION

Stream-water solute load estimation

Stream-water solute load, often referred to as mass flux, is the mass of chemical solutes or sediment transported at a point in a stream during a set period. Load estimation is frequently the central objective in both long-term research studies and water-quality monitoring programmes. In watershed studies, mass flux serves as an integrated measure of all processes within the watershed that affect water quality (Semkin *et al.*, 1994). With increased emphasis on watershed-based strategies for the control of non-point-source pollutants, reliable measures of loads over time are needed to address whether water quality is either improving or degrading. In the USA, stream reaches that do not meet water-quality standards are subject to waste-load allocation schemes based on the total maximum daily load (TMDL). A TMDL is defined as the maximum amount of a pollutant that a water body can receive and still meet water-quality standards (USEPA, 2000). Estimates of stream-water solute load are necessary to determine compliance with TMDLs.

---

* Correspondence to: Brent T. Aulenbach, US Geological Survey, Peachtree Business Center, 3039 Amwiler Road, Suite 130, Atlanta, GA, 30360-2824, USA. E-mail: btaulenb@usgs.gov
[†] Present address: CUAHSI, 2000 Florida Avenue, NW, Washington, DC 20009, USA.

The total solute load Φ_T is the convolution of solute concentration C and discharge Q over time t:

$$\Phi_T = \int C(t)Q(t) \, \mathrm{d}t \tag{1}$$

Load estimation using the integral in Equation (1) requires a continuous record of concentration and discharge. Although discharge can be measured quite readily, solute concentration typically is measured less frequently due to the expense of collecting and analysing samples for water quality. Therefore, concentration must be estimated between relatively infrequent samples. Several studies have compared different methods for load estimation (e.g. Dann *et al*., 1986; Preston *et al*., 1989). These methods can be categorized into four classes described below.

1. *Averaging methods.* All samples collected during an interval are averaged, either with or without weighting, to determine an average concentration for the interval. Load is estimated by multiplying this average concentration by the runoff for the same interval. If samples are not collected at a fixed interval or randomly through time, then concentrations should be flow weighted. The precision of the load estimate can be determined from the standard error of the average concentration.
2. *Period-weighted approaches.* Measured concentrations are assumed to represent a period around which the sample was collected, either as a constant value represented by a step function through time (e.g. Likens *et al*., 1977) or as a line connecting measured concentrations through time represented by a piecewise linear function (e.g. Larson *et al*., 1995). The product of concentration and discharge is summed through time to determine load. All samples collected are used in the estimation, and there is no preconceived model for concentration. The load estimate is sensitive to sampling frequency and distribution. Therefore, various sampling designs, such as selection at list time (SALT; Thomas, 1985) and time-stratified sampling (Thomas and Lewis, 1993), have been developed to reduce errors. The variance of the load estimate can be derived from a semivariogram calibrated to the data using a cross-validation technique (Shih *et al*., 1998).
3. *Regression-model (or rating-curve) methods.* $C(t)$ is estimated using a regression model relating concentration to a continuously measured variable such as discharge and day of year (e.g. Johnson, 1979; Cohn *et al*., 1992), thus enabling a direct calculation of Equation (1). Note that the variables used are often surrogates for underlying processes that partially control concentration. The load estimate is sensitive to the accuracy of the proposed model. A statistically based estimate of the precision of the load estimates can be determined (Gilroy *et al*., 1990). This method typically requires fewer data than the period-weighted approach and sometimes can be used beyond the period that samples were collected (Robertson and Roerish, 1999).
4. *Ratio estimators.* An average annual concentration is calculated by dividing the average daily load by the average daily discharge for days that have samples collected. Load is then determined by multiplying the average concentration by the runoff and by a factor to adjust for statistical bias due to non-representative sampling of the stream hydrograph (e.g. Beale, 1962). Ratio estimators assume that there is a linear relation through the origin between daily load and daily flow and that the variance around this line is proportional to the daily flow (Cochran, 1977).

The appropriate method to use depends on the frequency and distribution of sampling, the scale of the system, and the strength and form of the relation between concentration and other variables, such as flow and season (Richards and Holloway, 1987; Preston *et al*., 1989). Many studies have been done to assess errors in load estimates using various sampling designs and load estimation methods. For example, Preston *et al*. (1989) did a bootstrap (subsampling) study comparing load estimates with fixed-frequency sampling versus fixed-frequency and event sampling from averaging methods, regression-model methods, and ratio estimators. Guo *et al*. (2002) did a bootstrap study using various fixed-interval sampling frequencies to compare load estimates from various regression-model methods and ratio estimators at various time intervals. Robertson (2003) used various mixed-frequency sampling designs to assess errors in load estimates using the regression-model method.

Sampling design

Instrumentation advances, such as computerized automatic samplers and increased automation of laboratory analytical equipment, have made it feasible to collect many more samples to characterize a chemical time series. However, costs associated with collection and analysis of a sample are still considerable, and it remains impractical to collect water-quality data at uniform intervals at a sufficiently high frequency to characterize the short-term changes in concentrations that occur during storms. Therefore, chemical time series often consist of samples collected at a fixed interval (often weekly to monthly, depending on the size of the river basin) interspersed with high-frequency (on the order of minutes to hours) samples during storms, snowmelt, or events such as combined sewer overflows, giving rise to a mixed-frequency dataset. This type of sampling defines concentration patterns better through time during periods in which solute concentrations are more variable.

Purpose and scope

This paper presents the composite method as an alternative method for estimating solute loads. The composite method combines aspects of the regression-model method and the period-weighted approach and incorporates the use of mixed-frequency water-quality sampling. The bias and precision of the composite-method load estimates will be determined using data from Panola Mountain Research Watershed (PMRW) for different sampling scenarios and time intervals using a bootstrap experiment. The ability of the composite method to estimate load accurately and precisely will be assessed by comparing composite-method load estimates with load estimates from the period-weighted approach and regression-model method. Finally, we will demonstrate that the composite method can improve load estimates for shorter time intervals and allow for the better identification of patterns (trends) in the load estimates.

METHODS

Dataset description

This study uses data from PMRW, which has a long-term, comprehensive stream-water water-quality dataset that is necessary for the analysis. PMRW is a 41 ha experimental basin (Huntington *et al.*, 1993) located approximately 25 km southeast of Atlanta, Georgia, USA, in the southern Piedmont physiographic province. The stream draining PMRW is perennial at the watershed outlet. Streamflow is characterized by a sustained baseflow with rapid rise and fall during storms. Data used in this analysis spanned an 8-year period from water year (WY) 94 to WY01. (The US Geological Survey defines a WY as 1 October of the previous year to 30 September of the designated year; hence, WY94 is from 1 October 1993 to 30 September 1994.) Annual water yields for the 8-year period are variable, ranging between 16 and 46% (Figure 1).

Stream-water discharge was determined at 5 min intervals during baseflow conditions and at 1 min intervals during storms. To characterize the water quality of the stream, 1923 samples were collected during the study period. Grab samples were collected weekly, and storm-based samples were collected for the majority of large storms and some smaller storms using an automatic stage-activated sampler (large storms being defined herein as storms with peak flows greater than $10 \, l \, s^{-1}$ ($0 \cdot 21 \, cm \, day^{-1}$)). Storm-based samples were distributed in such a manner as to cover the range in flows, flow conditions, and dynamics of the streamflow hydrograph, with additional sampling during very high flows ($>100 \, l \, s^{-1}$, $2 \cdot 1 \, cm \, day^{-1}$). Hence, more samples were typically collected for larger and more complex, multipeaked storms than for smaller storms. Sampling distribution by WY is illustrated in Figure 1, with fewer storm samples collected in drier years, which contained fewer and smaller large storms.

Baseflow and storm periods were identified using a flow hydrograph criterion developed specifically for PMRW. Baseflow periods made up 56·6% of the total runoff during the study period (Table I) and daily average baseflow ranged from $0 \cdot 58 \, l \, s^{-1}$ ($0 \cdot 012 \, cm \, day^{-1}$) to $12 \cdot 3 \, l \, s^{-1}$ ($0 \cdot 26 \, cm \, day^{-1}$). During the study
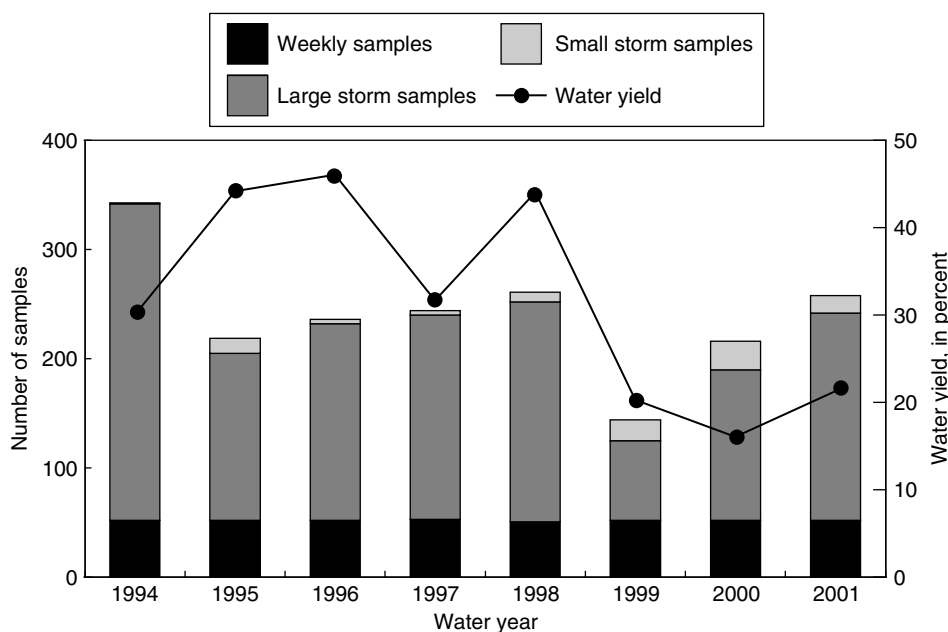
Figure 1. Number of weekly fixed-interval, large storm, and small storm samples collected each year at outlet of PMRW and annual water yields for the period WY94 to WY01

Table I. Summary of runoff and storm sampling by flow category.[a] It was assumed that all baseflow periods were sampled sufficiently

Flow category	Percentage of total runoff (%)	No. of samples	No. of storms	No. of storms sampled	Percentage of storms sampled (%)	Average no. of samples per sampled storm	Range of no. of samples per storm	Portion of runoff sampled in flow category (%)
Baseflow	56·6	354	n.a.	n.a.	n.a.	n.a.	n.a.	100·0
Small storm	4·4	121	219	80	36·5	1·5	1–4	45·7
Large storm	39·0	1448	232	178	76·7	8·1	1–36	90·7
Total	100·0	1923	n.a.	n.a.	n.a.	n.a.	n.a.	94·0

[a] n.a.: not applicable.

period, 232 large storms were identified, which contributed 39·0% of the total runoff. Of these large storms, 178 contained storm-based samples (76·7%), with an average of 8·1 samples collected per large storm and a range of from 1 to 36 samples collected per storm. The storms that were not sampled tended to be small, as 90·7% of the large storm runoff occurred during large storms containing samples. There were also 219 small storms identified, which contributed 4·4% of the total runoff. Of these small storms, 80 had samples collected during them (36·5%), averaging 1·5 samples collected per storm and ranging from one to four samples collected per storm. Small and large storms that were not sampled represent 6·0% of the total runoff for the study period.

Alkalinity and chloride analyses of the stream-water samples collected are considered for this study. Although both parameters undergo dilution with increasing flows, their controls on concentration variations differ somewhat at PMRW. Alkalinity decreases at higher flows are largely the result of the combination of dilution of the weathering products calcium and sodium present in deeper groundwater along with increases

in sulphate from flow paths through shallow soil horizons. Meanwhile, chloride concentrations decrease at higher flows as the result of dilution of higher chloride-concentrated water present in deep groundwater by increased contributions of water from shallower flow paths that have lower concentrations of chloride. Chloride in groundwater is derived from atmospheric deposition and is concentrated by evapotranspiration. Alkalinity and chloride are analysed using standard methods. Quality control included analysis of standard reference materials and routine interlaboratory comparisons. The coefficients of variation (COVs, i.e. the standard deviation divided by the mean, expressed as a percentage) for concentrations of standard reference materials were 2·2% for alkalinity and ranged from 2·8 to 5·5% for chloride. Further information on analytical techniques is given by Huntington *et al.* (1993).

Regression-model construction

The composite method, like other regression-model load-based approaches, uses a regression model in which concentration is a function of other continuous variables over time. In this study, concentration was estimated as a function of stream discharge and day of year. One model was constructed for each solute using all samples for the period of record. The concentration–discharge relation was modelled using a hyperbolic function (Johnson *et al.*, 1969), as its functional form fitted the data better than a log–log model. The hyperbolic function does not introduce the transformation bias that occurs for a log–log model because the hyperbolic function does not require retransformation from log space to linear space (Cohn *et al.*, 1989). The hyperbolic model also has some physical significance, in that it describes a two-component mixing model.

Average discharge for a certain period immediately preceding the time of sample collection was used instead of instantaneous discharge in the hyperbolic model, as suggested by Hooper (1986). The averaging period reduced the degree of hysteresis in the concentration–discharge relation. Hysteresis is the effect in which the concentration–discharge relation differs for samples collected during the rising limb of the hydrograph compared with during the falling limb of the hydrograph.

There was also a dummy variable for whether or not the sample was collected on the rising limb of a hydrograph. This dummy variable is 'on' when the instantaneous discharge is increasing and there has been precipitation in the previous hour. This term was added to the regression model to correct the problem that the means of the residuals were significantly different ($p < 0.05$) between the rising limb samples and all other samples.

Seasonal variations in concentration were modelled using sine and cosine functions. Periods of 1 year and 0·5 years were used for sine and cosine terms to fit asymmetrical annual cycles. Each pair of sine and cosine terms is equivalent to an individual sine function with the same period and a phase-shift term.

Therefore, our equation for model concentration is

$$
\begin{aligned}
C_M(t) = {} & C(Q^*(t), D) \\
= {} & a_0 + \frac{a_1}{1 + \beta Q^*(t)} + a_2 m + a_3 \sin\left(\frac{2\pi D}{365}\right) + a_4 \cos\left(\frac{2\pi D}{365}\right) \\
& + a_5 \sin\left(\frac{2\pi D}{182 \cdot 5}\right) + a_6 \cos\left(\frac{2\pi D}{182 \cdot 5}\right)
\end{aligned} \tag{2}
$$

where $Q^*(t)$ (l s^{-1}) is the average discharge before time t (depending on solute), m is a dummy variable depending on the slope of the hydrograph (0 or 1), D is the day of year (from 0 to 365), t is time, β is a fitted hyperbolic regression model parameter to linearize data, and a_n ($n = 0$–6) are regression model parameters.

The regression-model parameters a_n ($n = 0$–6) and β were determined using a non-linear calibration algorithm (SAS, 1990). Average discharges of 30 min for alkalinity and 15 min for chloride were used instead of instantaneous discharges, as the average discharges fit the data better. Residual concentrations are defined as the difference between the model-predicted and observed concentrations. Note that long-term time-trend

terms purposely were not included in the regression model. In the composite method, long-term trends are incorporated into the load estimates in a different manner, as discussed later.

It is important to assess the structure of the residuals to the model thoroughly. Residuals to any regression model should be statistically independent and identically distributed to ensure accurate model parameter estimation and model prediction. Unfortunately, mixed-frequency datasets have a tendency to violate these statistical assumptions (Hooper and Aulenbach, 1997). To assess compliance with statistical assumptions, residuals were plotted against the hyperbolic discharge term, day of year, and time.

Composite method for load estimation

The composite method combines aspects of the regression-model method and period-weighted approach to develop an alternative concentration function from which to estimate loads. One portion of the concentration function consists of the regression model developed above ($C_M(t)$), the same function used by the regression-model method. The second portion of the concentration function is a continuous function of regression-model residual concentrations ($C_\varepsilon(t)$) constructed by linearly interpolating the residuals through time in a piecewise manner between sample observations. The period-weighted approach often uses a similar approach, but for the observed concentrations. Although a piecewise linear function is a simplistic model of the residuals, and far more sophisticated models are possible, the data density typically available at research catchments make the approach plausible. Substituting the two concentration functions into the relation for integrating solute load over time (Equation (1)) yields

$$\Phi_T = \int C_M(t)Q(t)\,\mathrm{d}t - \int C_\varepsilon(t)Q(t)\,\mathrm{d}t \tag{3}$$

The different concentration functions used by the regression-model method, the period-weighted approach, and the composite method for a typical sampled storm hydrograph (Figure 2a) are illustrated in Figure 2b–d. Figure 2b shows $C_M(t)$ used by the regression-model and composite methods along with the piecewise linear function of the observed concentrations used by the period-weighted approach. Figure 2c shows the residuals to the regression model connected in a piecewise linear manner ($C_\varepsilon(t)$) used by the composite method. The combined concentration function for the composite method, $C_M(t) - C_\varepsilon(t)$, is shown in Figure 2d. Note that this concentration function predicts the observed concentration.

In this manner, the composite method attempts to reduce the error of load estimates by taking advantage of the strengths of both load estimation methods that compose it. Whereas the period-weighted approach varies concentrations between sample observations in either a stepwise or linear fashion, the composite method uses a regression model to predict typical concentration variations between sample observations. Whereas the regression-model method develops a model of average chemical response and then ignores the remaining information contained in the residual concentrations, the composite method retains the residual concentrations to adjust the regression-model-predicted concentrations to the observed sample concentrations and applies the residuals between samples using a period-weighted approach.

Therefore, the composite method total load Φ_T is composed of two components:

$$\Phi_T = \Phi_M - \Phi_\varepsilon \tag{4}$$

where the model load Φ_T is the load estimated from $C_M(t)$, and the residual load Φ_ε is the load estimated from $C_\varepsilon(t)$.

The Φ_M portion of Φ_T is equivalent to the load estimated using the regression-model method. The Φ_ε portion of Φ_T is the portion of the load unexplained by the regression model and represents a combination of unexplained variation due to biogeochemical and hydrological processes that are not accounted for in the regression model, to changes in the model relation over time (trend), and to measurement errors. The composite
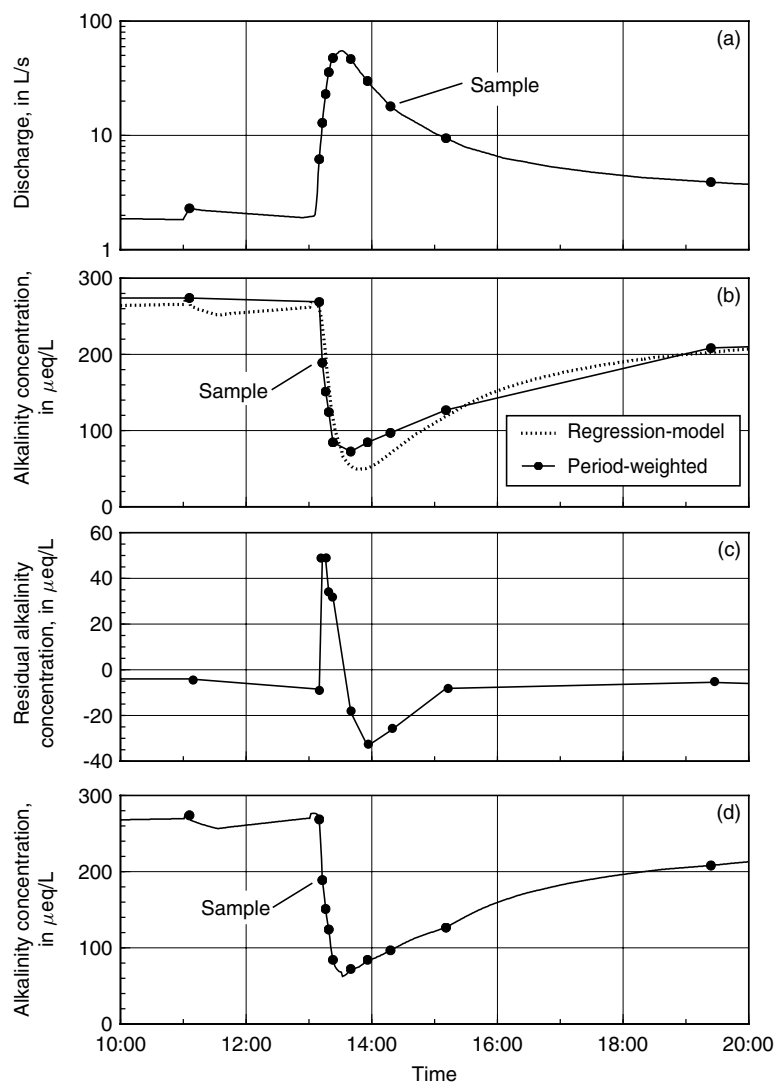
Figure 2. Concentration functions used in various load estimation methods for alkalinity at PMRW for storm on 7 August 1995. (a) Stream discharge at PMRW. Points indicate when samples were collected for water quality. (b) Period-weighted approach piecewise linear function $C_\varepsilon(t)$ and regression-model method function $C_M(t)$. Points indicate observed sample concentrations. (c) Piecewise linear function of residual concentrations $\varepsilon(t)$. Points indicate residual concentration when sample is collected. (d) Composite method concentration function. Note that $C(t)$ goes through observed concentrations (points)

method has been used in studies by Huntington *et al*. (1994) and Peters *et al*. (2006), and is equivalent to the Q-proportionate method employed by Vanni *et al*. (2001).

The integral for solute load (Equation (3)) is estimated numerically using the extended trapezoid rule (Press *et al*., 1986) with a convergence criterion of $<0.2\%$ change in load between iterations. Solute load was estimated on a daily time-step. A minimum of seven iterations (65 equally spaced evaluations per day of the integral in Equation (3), a calculation rate of one every 22·5 min) was used to ensure that storms with short duration entered the calculations. For each evaluation, the regression-model and residual concentrations along with the instantaneous discharge were determined and load was calculated. Convergence always occurred within 12 iterations (2049 evaluations per day, a calculation every 42 s).

Assessing error in estimated true load

The actual or 'true' load must be known in order to determine the error of load estimates accurately using various sample designs and methods in this study. Other load estimation methodology studies on larger streams for which loads could be sufficiently estimated on a daily time-step and for which daily concentrations were available have been able to calculate the true load directly (e.g. Preston *et al.*, 1989, Guo *et al.*, 2002). Smaller watersheds, such as PMRW, typically do not have water-quality measurements at a sufficient frequency to be able to calculate the true load directly and must, therefore, estimate the true load using the best sampling possible. Therefore, the estimated true load is the load estimated using the composite method that includes all samples collected (weekly fixed-interval sampling along with all small and large storm samples). Although sampling at PMRW is quite extensive, there is still the question of whether the sampling for the estimated true loads was at a sufficient frequency to determine the true loads accurately.

As the frequency of sampling is increased, the load estimates should converge on the true load (Preston *et al.*, 1989). Hence, if sampling is increased incrementally, then the incremental improvement in the load estimates should decrease as the estimate converges on the true load. Therefore, a small difference between loads estimated using a true-load sampling scenario and loads estimated from a subsampling scenario with significantly fewer samples is an indication that sampling was sufficient and that the estimated true load is accurate. A larger difference in load estimates is an indication that there was insufficient sampling and the estimated true load might be biased and imprecise. Unfortunately, these differences in load estimates only qualitatively assess the estimated true load, as these differences are indications of convergence and not a quantitative measure of the actual bias.

Based on the idea above, the estimated true load was assessed by comparison with load estimates from four subsampling scenarios to see whether the load estimates were close to converging on the true load. Each of the four subsampling scenarios represents a different aspect of sampling that requires the adequacy of sampling to be assessed. The adequacy of weekly fixed-interval sampling was assessed by comparing loads estimated using fixed-interval sampling reduced to biweekly. The adequacy of sampling enough of the large storms was assessed by comparing loads estimated with the subsampling scenario of 60% of large storms sampled along with weekly fixed-interval sampling (a test case from the bootstrap experiment described below). The adequacy of sampling the large storms at a sufficient frequency was assessed by comparing loads estimated with a subsampling scenario that included all large storms sampled, but included only every second storm sample collected. The importance of sampling small storms was assessed by comparing loads estimated with a subsampling scenario that did not include the small storm samples.

A rough estimate of the affect of unsampled small and large storms on the overall precision of the load estimates was determined by the product of the percentage of the unsampled storm runoff (with respect to the total runoff) and the COV of the residuals (COVs of residuals are calculated herein as the standard deviation of the residuals divided by the mean of the observed concentrations and are expressed as a percentage). This estimate assumes that the COVs of the residuals are statistically independent and identically distributed.

Regression-model method

Regression-model method loads were estimated to compare the results with composite-model load estimates. Regression-model method loads were estimated using the same regression model developed for the composite method. Regression-model method loads were also estimated using regression models with additional, often used, quadratic long-term trend terms (time and time-squared terms) to determine how these terms affect the load estimates.

Bootstrap experiment

A bootstrap (subsampling) experiment using the PMRW dataset was designed to determine and compare the bias and precision of the load estimates for the composite method and the period-weighted approach using a variety of different sampling designs. A bootstrap experiment was chosen rather than using a Monte Carlo

approach with an artificial dataset to illustrate how the composite method can incorporate natural variations in the dataset that were not explained by the regression model. A Monte Carlo approach has its own advantages, in that the true load is known and the variability in the underlying data is known and random, allowing for the effects of variation on load estimates to be more easily assessed. This is because a Monte Carlo approach uses an artificial dataset with a known function of concentration through time with a known introduced error. Artificial datasets, however, cannot easily mimic variations observed in real datasets.

The effect of varying numbers of large storms sampled on load estimates was simulated by setting the percentage of large storms sampled during the study period to 0, 10, 20, 40, 60, and 77% (with 77% representing the total percentage of large storms sampled during the study). The effect of the frequency of routine fixed-interval sampling on load estimates was simulated using two different fixed intervals, i.e. weekly and monthly. Hence, there are 12 combinations of percentage of storm sample inclusion and fixed-interval sampling frequencies that were used to estimate loads. These combinations are referred to hereafter as test cases. For test cases with between 10 and 60% large storms, load estimates were simulated 100 times, with the large storms sampled being randomly selected for each simulation (with all samples from each selected storm included in the simulation). In this manner, the variability in the load estimates due to the random selection of storms is also assessed for these test cases. All test cases were estimated using both the composite method and the period-weighted approach.

Bias and precision of load estimates

Load estimates from the subsampling scenarios, bootstrap experiment, and regression-model method were compiled monthly, quarterly, annually (WY), and for the 8-year study period. The error in load estimates is a combination of bias (accuracy) and precision. Bias and precision were calculated as the percentage of the estimated true loads to allow easy comparisons between the two solutes and with other studies, and to give equal weight to each month, quarter, or year when calculating precisions for different time intervals. The bias of each test case is calculated as the percentage difference of the load estimate with respect to the estimated true load. The average load estimate was used for comparison in bootstrap experiment test cases that contained 100 simulations. The precision of each test case was determined at the various time-scales as the standard deviation of the percentage errors. The precision is a measure of the variability in the percentage errors of the load estimates at shorter time intervals and the variability in the load estimates due the random selection of large storms. Note that precision for the study period time-scale could only be determined for the bootstrap experiment test cases with between 10 and 60% large storms, as these test cases had 100 simulations and allowed for the calculation of a standard deviation.

Note that the biases and precisions determined for the load estimates only assess errors as the result of the load estimation technique and sampling scenarios with respect to the estimated true loads. Additional errors associated with flow measurements, the representativeness of the water-quality samples collected, and chemical analyses are not considered herein, but each can typically be on the order of 5 to 10%. The cumulative effect of all these errors should be considered when determining the overall error in the load estimates.

Trend analysis approach

The composite method is formulated such that the effects of hydrologic variability on load estimates related to regression-modelled relations are incorporated into Φ_M. The concentration deviations from the regression model are incorporated into Φ_ε. A runoff-normalized Φ_ε, which represents a flow-weighted average concentration deviation from the model-predicted concentrations, can be calculated for a certain period (e.g. year, quarter, or month) by dividing by the runoff for that period. The runoff-normalized Φ_ε can be used to identify temporal patterns in concentration deviations from the regression model and can be used to investigate other more long-term controls on concentrations not modelled by the regression model. Relations between runoff-normalized Φ_ε and water yields were explored at various time-steps.

RESULTS

Evaluation of regression models

Model performance. The regression models developed explained a large portion of the variation in concentrations, with model R^2 values of 0·93 for alkalinity and 0·80 for chloride (Table II). It should be noted that, despite the higher model R^2 value for alkalinity, alkalinity still had a higher COV in the residuals of 16·7% compared with 12·1% for chloride. This occurred because the COVs of observed concentrations were much higher for alkalinity (62·8%) than for chloride (26·8%).

Residual analysis. There were persistent long-term patterns in the baseflow sample residuals (lack of serial independence), as illustrated in Figure 3 for alkalinity. There were also larger short-term variations in the storm sample residuals compared with baseflow samples (unequal variance), as illustrated in Figure 4 for alkalinity in WY95. This occurred despite the fact that the overall residual patterns for baseflow and storm samples were homoscedastic when plotted versus transformed discharge and day of year. These observations indicate that the models have some degree of misspecification, neither predicting the long-term patterns in concentrations during baseflow conditions nor adequately capturing the short-term variations in concentrations during storms.

Assessment of load estimation methods

Composite method load estimates. The portion of the total load determined from the regression model Φ_M is always the dominant component of the load, representing 97·9% of Φ_T for alkalinity and 99·4% of Φ_T for chloride during the study period (Table III). This was expected, because the regression models used in this analysis explained a large portion of the variability in the solute concentrations. Φ_ε as a percentage of the total load was −2·1% for alkalinity and −0·6% for chloride for the study period. The Φ_ε contributions are

Table II. Summary of regression model R^2 values and COVs of solute concentrations and residuals at PMRW for period of study. The mean of the concentrations is used to calculate the COV for the residuals

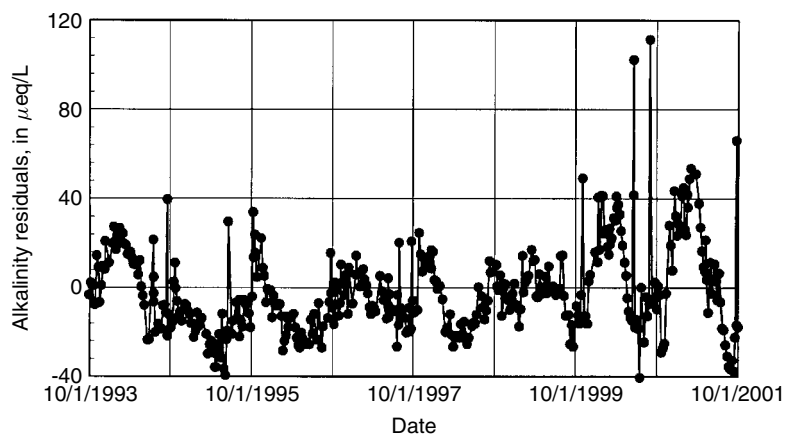| Solute	Model R^2	COV of concentrations (%)	COV of residuals (%)
Alkalinity	0·93	62·8	16·7
Chloride	0·80	26·8	12·1

Figure 3. Baseflow sample residual concentrations for alkalinity at PMRW for period of study, WY94−WY01
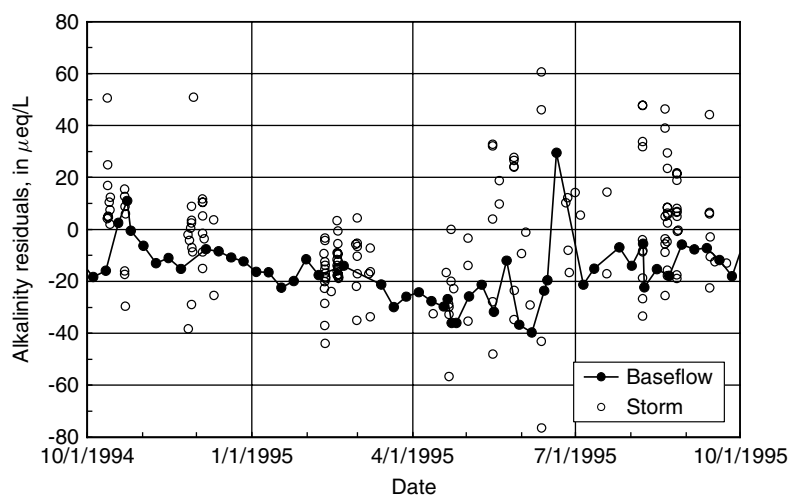
Figure 4. Alkalinity residual concentrations for baseflow and storm samples at PMRW for WY95

Table III. Comparison of model load Φ_M and residual load Φ_ε to total load Φ_T estimates from the composite method at PMRW for period of study and annually

Solute	Period of study Φ_M/Φ_T (%)	Period of study Φ_ε/Φ_T (%)	Minimum annual Φ_ε/Φ_T (%)	Maximum annual Φ_ε/Φ_T (%)
Alkalinity	97·9	−2·1	−7·8	8·8
Chloride	99·4	−0·6	−3·7	5·4

fairly close to zero since the residuals concentrations vary around a mean of zero; therefore, Φ_ε values tend to just about cancel out over time. But on an annual basis, Φ_ε typically represents a much larger portion of the total loads, with the range in the annual percentage contribution of Φ_ε to Φ_T being from −7·8% to +8·8% for alkalinity and −3·7% to +5·4% for chloride. The large range in annual Φ_ε is an indication of persistent (on an annual basis) yet variable (for the entire study period) deviations in observed concentrations from the predicted concentrations of the regression models.

Assessment of estimate of true loads. The bias and precision in the subsampling scenario load estimates with respect to the estimated true loads are summarized in Table IV. Subsampling scenarios resulted in a ±0·3% or less bias for the study period. These small differences in load estimates indicate that the sampling used in the estimated true loads is sufficient for determining an unbiased estimate of load for the study period. The precision of the subsampling scenario load estimates with respect to the estimated true loads increased as one decreased the time period of interest, getting as high as 2·1% for alkalinity and 1·4% for chloride on a monthly basis. These observed variations indicate that the adequacy of the sampling of these subsampling scenarios is somewhat insufficient at these shorter time intervals. This suggests that there may be biases in the estimated true loads at these shorter time intervals that could result in errors in the precision estimated for test cases at shorter time intervals.

The effect on precision of the load estimates from the unsampled large and small storms portion of runoff, which accounted for 6·0% of the total runoff of the study period, was estimated as 1·0% for alkalinity and 0·7% for chloride. These precision estimates may not be accurate because they assume that the residual concentrations are independent and identically distributed, whereas it has been already shown that they are not.

Table IV. Bias and precision of load estimates from subsampling scenarios with respect to the estimated true estimates for various time-scales

Subsampling scenario[a]	Bias for study period (%)		Standard deviation					
	Alkalinity	Chloride	Annual percentage errors (%)		Quarterly percentage errors (%)		Monthly percentage errors (%)	
			Alkalinity	Chloride	Alkalinity	Chloride	Alkalinity	Chloride
Biweekly fixed interval ($n = 1778$)	−0·1	0·2	0·6	0·3	1·1	0·8	1·7	1·4
60% large storms ($n = 1514$ on average)	0·2	−0·3	0·8	0·5	1·5	1·0	2·1	1·4
Every other large storm sample ($n = 1218$)	0·3	0·1	0·4	0·5	1·2	0·8	1·8	1·1
Exclude small storm samples ($n = 1830$)	0·2	−0·1	0·2	0·1	0·5	0·3	0·8	0·5

[a] n is the number of samples.

Biases of load estimates. Biases in load estimates from the bootstrap experiment and from the regression-model method are summarized for the study period in Figure 5. Composite method load estimates are fairly unbiased and had fairly similar accuracies for both solutes, with biases ranging from −0·8% to +0·7% for alkalinity and −0·8% to +0·5% for chloride. For weekly fixed-interval sampling test cases, biases decreased as the percentage of large storms sampled was increased from 0 to 77%. But load estimates are somewhat biased when using monthly fixed-interval sampling, with errors for the all-large storms test case of −0·8% for alkalinity and 0·5% for chloride. The magnitude of the bias was equivalent to the bias obtained by the test case using weekly fixed-interval sampling with no large storm sampling.

In test cases in which 77% of the large storms were included, the period-weighted approach underestimated loads. Load estimates were fairly unbiased when weekly fixed-interval sampling was used, being −1·4% for alkalinity and −0·5% for chloride, but biases were larger when monthly fixed-interval sampling was used, these being −7·6% for alkalinity and −1·9% for chloride (Figure 5). These biases were likely an artefact of storm sampling design. The first storm sample was not taken until after the initial rise in streamflow, and the last storm sample was taken before flow and concentration completely returned to baseflow conditions. As a consequence, storm samples with lower concentrations near the beginning and end of each storm were applied to the adjacent baseflow periods, resulting in load being underestimated during these baseflow periods. This is more evident for the monthly fixed-interval sampling test case, where there were fewer baseflow samples collected.

The period-weighted approach overestimated loads when 60% or fewer large storms were sampled (Figure 5). Overestimates are considerable in test cases for which 40% or fewer large storms were included, with errors ranging from 5·7 to 23% for alkalinity and from 3·4 to 10% for chloride. This bias is the result of higher baseflow concentrations being applied to storm periods, which actually have lower concentrations due to dilution, due to the lack of storm sampling. Overall, the chloride load estimates are less biased than the alkalinity load estimates, largely because chloride sample concentrations were less variable than alkalinities (Table II).

Errors in regression-model method load estimates, using the regression models used in the composite method (with no long-term time terms), are −2·1% for alkalinity and −0·6% for chloride. In comparison with the composite-method load estimates, the regression-model method errors are somewhat larger for alkalinity and similar for chloride. Adding long-term time and time-squared terms to the regression model, terms typically included in the regression-model method, reduced the bias to −1·1% for alkalinity and −0·1% for chloride.
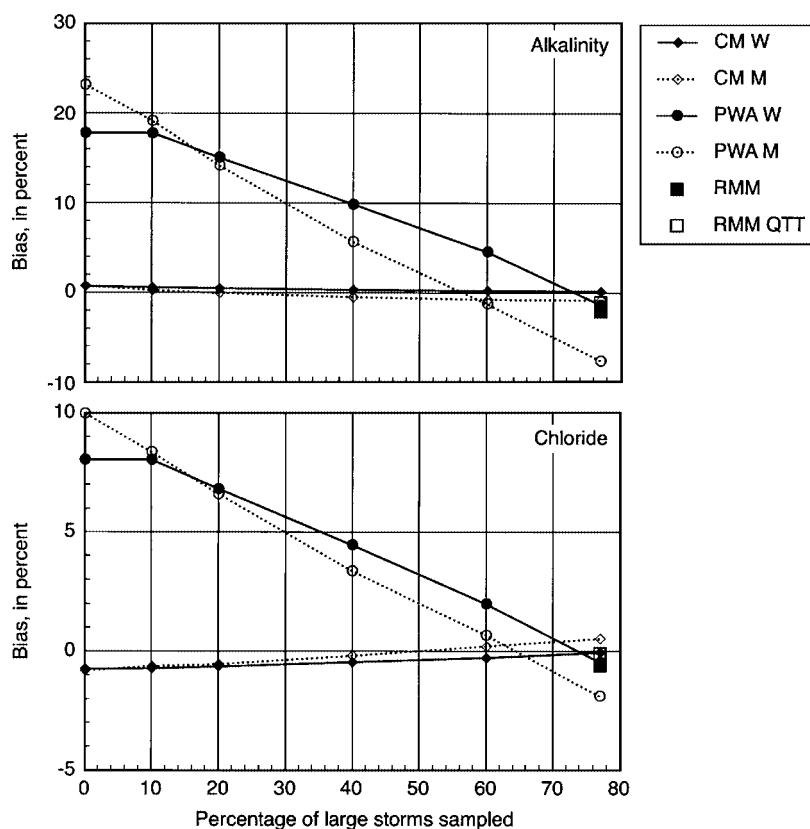
Figure 5. Biases in load estimates, with respect to the estimated true loads, using various load estimation methods and sampling designs for alkalinity and chloride for the study period (CM: composite method; PWA: period-weighted approach; RMM: regression-model method; W: weekly fixed-interval sampling; M: monthly fixed-interval sampling; QTT: regression model contains quadratic long-term time terms)

Temporal patterns in errors are illustrated for a select group of load estimation methods and sampling test cases on an annual basis for alkalinity in Figure 6. Annual errors for the composite-method load estimates for the test case with 77% of large storms sampled and weekly fixed-interval sampling are always within $\pm 0.5\%$. With some exceptions, the annual errors are generally less than $\pm 2\%$ for other composite-method test cases. The period-weighted approach using the most ideal sampling from the bootstrap experiment generally has annual errors of less than $\pm 2\%$, with the exceptions of WY00 and WY01, which have larger errors. Annual errors in load estimates using the regression-model method are much more variable, despite there being only small biases in the overall loads. Annual errors are as large as 8.8% when no time terms are included in the regression model, whereas errors are somewhat smaller when the regression model contained quadratic time terms, with errors of less than $\pm 6.1\%$. The addition of the quadratic time terms did not remove the overall pattern of annual deviations from the estimated true loads. Similar patterns in annual errors are observed for chloride, but with overall errors being about 20% lower.

Precision of load estimates. Precisions for the various load approaches, bootstrap sampling designs, and time intervals are summarized for alkalinity for selected test cases in Figure 7. As one would expect, load estimates are less precise as the time interval of interest is shortened and the amount of sampling is decreased. For the composite method, precisions for all test cases range from 0.2 to 0.5% for the study period, from 0.2 to 1.9% on an annual basis, from 0.5 to 4.0% quarterly, and from 0.8 to 5.6% monthly. Precision is better for test cases that had more samples collected, typically with somewhat better precisions obtained for the
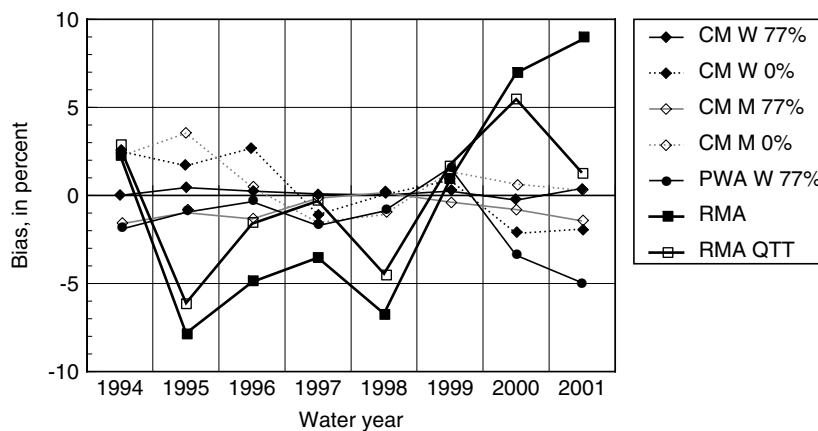
Figure 6. Biases in alkalinity annual load estimates, with respect to the estimated true load, using selected load estimation methods and sampling designs (CM: composite method; PWA: period-weighted approach; RMM: regression-model method; W: weekly fixed-interval sampling; M: monthly fixed-interval sampling; QTT: regression model contains quadratic long-term time terms; percentages indicate the percentage of large storms sampled)
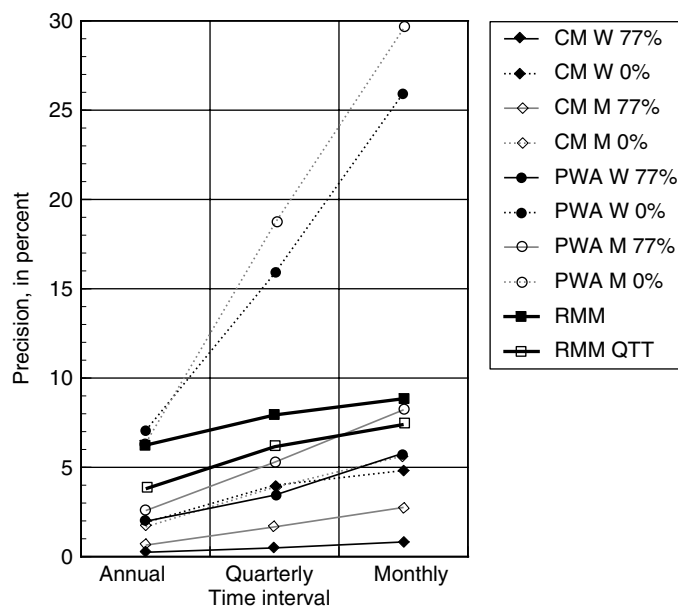

Figure 7. Precisions in alkalinity load estimates using selected load estimation methods, sampling designs, and time intervals (CM: composite method; PWA: period-weighted approach; RMM: regression-model method; W: weekly fixed-interval sampling; M: monthly fixed-interval sampling; QTT: regression model contains quadratic long-term time terms; percentages indicate the percentage of large storms sampled)

weekly fixed-interval test cases when compared with monthly fixed-interval test cases with about the same total number of samples collected during the period.

For the period-weighted approach, precisions for alkalinity are much more variable than the composite method, ranging from 1·2 to 2·1% for the study period, from 2·0 to 7·1% on an annual basis, from 3·4 to 19% quarterly, and from 5·8 to 30% monthly (Figure 7). The higher precisions observed for the period-weighted approach are to be expected; this is because the COVs in the observed concentrations, used in the period-weighted approach, are significantly higher than the COVs in the residual concentrations, which are
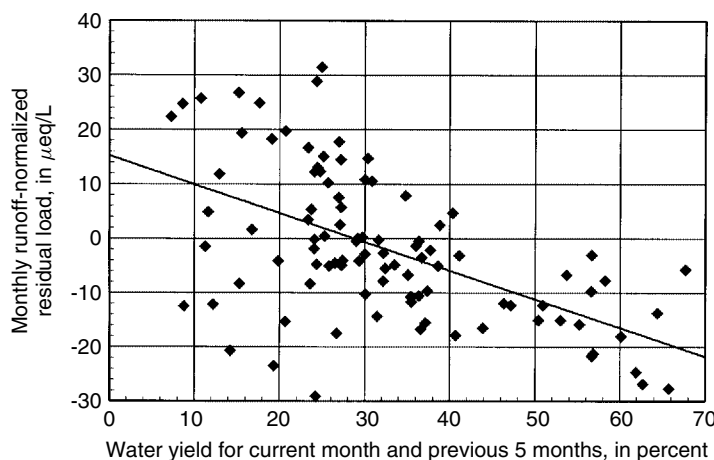
Figure 8. Water yield for period of previous 5 months and current month versus runoff-normalized Φ_ε for alkalinity. Linear trend-line has R^2 of 0·31

period-weighted in the composite method (Table II). The variability of the period-weighted approach for the test case with the most number of samples included is equivalent to the variability with the composite method for the test case with the least number of samples included.

Regression-model method precisions for alkalinity models with and without quadratic time terms are 3·8% and 6·3% respectively on an annual basis, 6·1% and 7·9% respectively quarterly, and 7·4% and 8·8% respectively monthly (Figure 7). Although the addition of the quadratic time terms improved the precision, regression-model method load estimates are more variable than all composite-method load estimates and are more variable than some of the better-sampled test cases using the period-weighted approach, especially at the annual time interval. The regression-model method, even with the inclusion of time terms, just does not have ability to model loads precisely during shorter time intervals. Chloride estimates have similar patterns in variability (not shown), but are more precise, with precisions being about 30% smaller for the composite- and regression-model methods and about 50% smaller for the period-weighted approach.

Patterns in residual loads

Annual runoff-normalized Φ_ε for alkalinity decreased linearly with annual water yield ($R^2 = 0.78$), indicating that the regression model overpredicted alkalinities during dry years and underpredicted alkalinities during wet years for the same flow and season conditions. Monthly runoff-normalized Φ_ε values for alkalinity were most strongly linearly related to the water yield of the period that included the current month and previous 5 months ($R^2 = 0.31$; Figure 8). Chloride runoff-normalized Φ_ε values did not have any significant relations with water yield.

DISCUSSION

Although the overall biases for the study period were always ±0·8% or less for all test cases using the composite method, the bootstrap experiment indicated that test cases with monthly fixed-interval sampling were often somewhat biased. Load estimates did not converge on the estimated true load when the number of large storms sampled was increased. There are two reasons why this occurred. First, when monthly fixed-interval sampling was used, baseflow periods were more often influenced by the residual concentrations of the first and last samples of a storm, which were not always representative of baseflow residual concentrations. Second, there are systematic patterns in the baseflow residuals that were accounted for better in the load estimates when

the higher weekly frequency fixed-interval sampling (of which about 85% represented baseflow conditions) was used. Still, the resulting biases for the monthly fixed-interval sampling test cases were always small and were either similar to or smaller than the magnitude of the overall bias observed in the regression-model method load estimates.

Although the regression-model method was adequate for accurately estimating loads for the overall study period, the method could not precisely estimate loads for shorter time periods, from annually to monthly. The regression model predicts the average concentration response for the study period with the premise that the model residuals represent only error. If this were the case, then residuals should vary randomly through time and residual loads should tend to cancel out over time, resulting in little error in the load estimates at these shorter time intervals. If this were the case, the composite-method load estimates would be no better than loads estimated using the regression-model method. But the residual analysis showed that the residuals do not vary randomly through time, indicating that the regression model failed to explain some of the variability in the observed concentrations. The addition of long-term quadratic time trend terms in the regression model only minimally improved load estimates for shorter time periods because the patterns of the deviations from the regression model did not follow a simple quadratic mathematical formulation.

Meanwhile, the composite method is better able to estimate loads at shorter time intervals because of its ability to incorporate systematic deviations in concentrations from the regression model into the load estimates. The composite method does this by retaining the information contained in the residuals to adjust the regression-model-predicted concentrations to those of the observed concentrations. This does not require a preconceived mathematical formulation of the patterns in the residual concentrations.

The relations observed for alkalinity runoff-normalized Φ_ε values are another indication that the residuals contain more than random error. Alkalinity concentration variations are not only a function of flow, but also of wetness conditions. The relation between the monthly runoff-normalized Φ_ε values and the water yields for the period of the current month and preceding 5 months indicate that alkalinity is not only a function of the current wetness conditions, but also is a result of how wet it has been lately. This alludes to a watershed process of retention and release. In fact, Huntington *et al*. (1994) showed that the export of sulphate to stream water at PMRW is increased after extended dry periods. This would result in a corresponding decrease in alkalinity during these conditions, which is what was observed. The composite method, as a result of its two-component formulation of load, thereby provides a useful mechanism for exploring unmodelled relations between concentration and hydrologic conditions and/or biogeochemical processes, which can result in a better understanding of watershed processes.

The composite method should be applicable to other solutes and to watersheds of all sizes and should generally result in either improved or similar load estimates compared with loads estimated using the period-weighted approach or regression-model method. Any solute for which a regression model can be developed to predict concentration variations can be used in the composite method. For solutes that have regression models with little predictive ability, however, the regression model will then, in essence, just predict the mean concentration of the samples and the composite model will be equivalent to a period-weighted approach. The composite method should improve load estimates compared with the regression-model method if serial autocorrelation exists in the residual concentrations, as discussed previously. But non-ideal sampling with respect to patterns in the residual concentration could result in biases in the composite method, as observed for the monthly fixed-interval sampling test cases. Although the resulting biases were small in this study, biases could be conceivably more significant for a particular combination of sampling designs and patterns in residual concentrations; possibly resulting in the degradation of load estimates with respect to the regression-model method.

Sample frequency requirements for the composite method to obtain a desired accuracy and precision will vary by both watershed and solute characteristics. Larger watersheds, in which variations in flows and concentrations occur at a much lower frequency, typically would require less sampling than was needed for PMRW. Higher variability in residual concentrations will reduce the precision of the load estimates. Patterns in the residual concentrations through time can result in less precise load estimates at shorter time periods and

slightly biased load estimates overall unless adequate sampling is used to capture the patterns in the residuals adequately.

Further reductions in data requirements for load estimation methods will require models that more accurately capture the processes controlling concentration variation and more sophisticated statistical techniques that can better reproduce the statistical behaviour of model residuals. Sampling always will be necessary to estimate loads accurately during shorter time intervals, if concentrations systematically deviate from concentration-model predictions for periods that appreciably contribute to the overall load for the time period of interest.

CONCLUSIONS

Although the regression models developed for PMRW data could explain much of the variation in concentrations of solutes, the residual analysis indicated that residual concentrations behave differently during baseflow versus storm conditions (heteroscedasticity). The regression models developed did not account for short-term variations in concentration during storms and longer-term persistent concentration patterns observed during baseflow. These model misspecifications can result in errors in load estimates when they are based on regression models.

Period-weighted approach load estimates are very sensitive to the sampling design and typically result in large biases in load estimates when there is not sufficient sampling to capture the concentration variations during storms. Precisions of the load estimates were typically high for shorter time intervals of quarterly and monthly. However, the period-weighted approach did have low error with extensive sampling. For example, the comprehensive PMRW dataset used in the bootstrap experiment indicated that, when all samples were included, biases of less than $\pm 1.5\%$ were obtained for load estimates for the period of study and precisions of $<2\%$ were obtained on an annual basis.

The regression-model method was adequate for accurately estimating load for the 8-year study period, with biases ranging from -0.1 to -2.1% for the two solutes (alkalinity and chloride) and two model forms (no time terms versus quadratic time terms). The method had shortcomings when estimating loads for shorter time periods. This is because the model predicted the average concentration response for the conditions, not the specific response observed for a particular time period, resulting in poor precisions at time intervals from annually to monthly. Although the addition of quadratic long-term time terms in the regression model improved the overall bias in the load estimates, along with some improvements in precision at shorter time intervals, the time terms could not fully capture short-term concentration deviations from the regression model.

The composite method combines aspects of two commonly used methods: the regression-model method (which is used to predict variations in concentrations between collected samples) and a period-weighted approach (which is used to apply the residual concentrations from the regression model over time). The composite method is better able to adapt to short-term persistent deviations from the regression model observed in the PMRW data while being less sensitive to sampling deficiencies than the period-weighted method. The result is that the composite method load estimates are less biased and more precise than load estimates from the regression-model method and the period-weighted approach with no additional sampling requirements. The composite method should be applicable to other solutes for which a regression model can be developed for concentration versus other continuous variables and should be applicable to watersheds with different scales.

The bootstrap experiment using data from PMRW indicated that composite-method load estimate biases were always $\pm 0.8\%$ or less, even for test cases with relatively infrequent sampling. Precisions were always $<2\%$ on a study period and annual basis, and $<2\%$ precisions could be obtained for quarterly and monthly time intervals for test cases with better sampling. Ultimately, the bias and precisions in load estimates obtained with the composite method are dependent on the amount of variation in residual concentrations, the behaviour of residual concentrations through time, and the sampling design. The errors achieved in

composite-method load estimates with PMRW data are well within the typical range of errors associated with sample representativeness, flow measurement, and analytical chemistry.

The Φ_ε portion of the composite-method load estimate, runoff normalized, is useful for exploring patterns (trends) contained in the unmodelled portion of the load. At PMRW, monthly runoff-normalized Φ_ε values for alkalinity indicated that systematic deviations in alkalinity from the regression-model-predicted concentrations were related to recent wetness conditions.

REFERENCES

Beale EML. 1962. Some uses of computers in operational research. *Industrielle Organisation* **31**: 51–52.
Cochran WG. 1977. *Sampling Techniques*, 3rd edn. Wiley: New York.
Cohn TA, DeLong LL, Gilroy EJ, Hirsch RM, Wells DK. 1989. Estimating constituent loads. *Water Resources Research* **25**(5): 937–942.
Cohn TA, Caulder DL, Gilroy EJ, Zynjuk LD, Summers RM. 1992. The validity of a simple statistical model for estimating fluvial constituent loads: an empirical study involving nutrient loads entering Chesapeake Bay. *Water Resources Research* **28**(9): 2353–2363.
Dann MS, Lynch JA, Corbett ES. 1986. Comparison of methods for estimating sulfate export from a forested watershed. *Journal of Environmental Quality* **15**(2): 140–145.
Gilroy EJ, Hirsch RM, Cohn TA. 1990. Mean square error of regression-based constituent transport estimates. *Water Resources Research* **26**(9): 2069–2077.
Guo Y, Markus M, Demissie M. 2002. Uncertainty of nitrate-N load computations for agricultural watersheds. *Water Resources Research* **38**(10): 3-1–3-12.
Hooper RP. 1986. *The chemical response of an acid-sensitive headwater stream to snowmelt and storm events: a field study and simulation model*. PhD dissertation, Cornell University, Ithaca, NY (unpublished).
Hooper RP, Aulenbach BT. 1997. Trend analysis of mixed-frequency water-quality time series [Abs.]. *EOS, Transactions of the American Geophysical Union* **78**(46): F327.
Huntington TG, Hooper RP, Peters NE, Bullen TD, Kendall C. 1993. *Water, energy, and biogeochemical budgets investigation at Panola Mountain Research Watershed, Stockbridge, Georgia—a research plan*. Open-File Report 93–55, US Geological Survey, Denver, CO.
Huntington TG, Hooper RP, Aulenbach BT. 1994. Hydrologic processes controlling sulfate mobility in a small forested watershed. *Water Resources Research* **30**(2): 283–295.
Johnson AH. 1979. Estimating solute transport in streams from grab samples. *Water Resources Research* **15**(5): 1224–1228.
Johnson NM, Likens GE, Bormann FH, Fisher DW, Pierce RS. 1969. A working model for the variation in streamwater chemistry at the Hubbard Brook Experimental Forest, New Hampshire. *Water Resources Research* **5**(6): 1353–1363.
Larson SJ, Capel PD, Goolsby DA, Zaugg SD, Sandstrom MW. 1995. Relations between pesticide use and riverine flux in the Mississippi River basin. *Chemosphere* **31**(5): 3305–3321.
Likens GE, Bormann FH, Pierce RS, Eaton JS, Johnson NM. 1977. *Biogeochemistry of a Forested Ecosystem*. Springer-Verlag: New York; 43–44, 54–55.
Peters NE, Shanley JB, Aulenbach BT, Webb RM, Campbell DH, Hunt R, Larsen MC, Stallard RF, Troester JW, Walker JF. 2006. Water and solute mass balance of five small, relatively undisturbed watersheds in the U.S. *Science of the Total Environment* **358**: 221–242.
Press WH, Flannery BP, Tuekolsky SA, Vetterling WT. 1986. *Numerical Recipes: The Art of Scientific Computing*. Cambridge University Press: New York; 111–112.
Preston SE, Bierman VJ Jr, Silliman SE. 1989. An evaluation of methods for the estimation of tributary mass loads. *Water Resources Research* **25**(10): 1379–1389.
Robertson DM. 2003. Influence of different temporal sampling strategies on estimating total phosphorus and suspended sediment concentration and transport in small streams. *Journal of the American Water Resources Association* **39**(5): 1281–1308.
Robertson DM, Roerish ED. 1999. Influence of various water quality sampling strategies on load estimates for small streams. *Water Resources Research* **35**(12): 3747–3759.
Richards RP, Holloway J. 1987. Monte Carlo studies of sampling strategies for estimating tributary loads. *Water Resources Research* **23**(10): 1939–1948.

SAS. 1990. *SAS/STAT User's Guide*, Volume 2. SAS Institute, Inc.: Cary, NC; 1135–1193.

Semkin RG, Jefferies DS, Clair TA. 1994. Hydrochemical methods and relationships for study of stream output from small catchments. In *Biogeochemistry of Small Catchments: A Tool for Environmental Research*, Moldan B, Cerny J (eds). Wiley: New York; 163–187.

Shih G, Wang X, Grimshaw HJ, VanArman J. 1998. Variance of load estimates derived by piecewise linear interpolation. *Journal of Environmental Engineering* **124**(11): 1114–1120.

Thomas RB. 1985. Estimating total suspended sediment yield with probability sampling. *Water Resources Research* **21**(9): 1381–1388.

Thomas RB, Lewis J. 1993. A comparison of selection at list time and time-stratified sampling for estimating suspended sediment loads. *Water Resources Research* **29**(4): 1247–1256.

USEPA. 2000. *Overview of current total maximum daily load—TMDL—program and regulations*. EPA841-F-00-009, US Environmental Protection Agency, Washington, DC.

Vanni MJ, Renwick WH, Headworth JL, Auch JD, Schaus MH. 2001. Dissolved and particulate nutrient flux from three adjacent agricultural watersheds: a five-year study. *Biogeochemistry* **54**: 85–114.